View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by VU Research Portal

VU UNIVERSITEIT AMSTERDAM

VU Research Portal

An approximation to the distribution of quadratic forms in many normal variables

Sneek, J.M.; Smits, J.

1990

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Sneek, J. M., & Smits, J. (1990). *An approximation to the distribution of quadratic forms in many normal variables*. (Serie Research Memoranda; No. 1990-49). Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 22. May. 2021

1990-49

Faculteit der Economische Wetenschappen en Econometrie

ET

05348

Serie Research Memoranda

An approximation to the distribution of quadratic forms in many normal variables

J.M. Sneek
J. Smits

Research Memorandum 1990-49
October 1990

vrije Universiteit *amsterdam*

BIBL. VRIJE UNIVERSITEIT
AMSTERDAM
ET.
05348.-

AN APPROXIMATION TO THE DISTRIBUTION OF QUADRATIC FORMS IN MANY NORMAL VARIABLES

by

J.M. Sneek
J. Smits

Free University of Amsterdam
Faculty of Economics and Econometrics
De Boelelaan 1105
1081 HV Amsterdam

Summary

In this paper we seek to approximate to the distribution of general large quadratic forms in normal variables in at most $\mathcal{O}(n^2)$ arithmetic operations. The main idea is to split the quadratic form in a part containing the dominant eigenvalues and a remaining part that is approximated using the normal distribution. The dominant eigenvalues are obtained through a generalized power method, thus avoiding the expensive job of finding all eigenvalues. In special cases the method may involove only $\mathcal{O}(n)$ operations.

1. INTRODUCTION

In this paper the problem of approximating the probability $P(Q \leq c)$ is dealt with, where Q is a quadratic form in normal variables

$$Q = x'Ax, \tag{1}$$

A being a (non-stochastic) symmetric $n \times n$ matrix and x a normally distributed vector with unit covariance matrix. Here we only consider the special but important case that the expectation of x is zero.

It should be noted that tail probabilities of ratios of quadratic forms can be rewritten as tail probabilities of quadratic forms and both arise in many statistical applications, e.g. in tests for serial correlation in regression models (see e.g. FAREBROTHER (1980)), in significance tests in regression models (see e.g. KIVIET (1980) and PALM and SNEEK (1984)), tests for unit roots (see e.g. EVANS and SAVIN (1981, 1984)), in ARMA models when the distribution of autocorrelations is considered (see e.g. SNEEK (1983), ALI (1984) and DUFOUR and ROY (1985) and in certain point-optimal-invariant tests (see e.g. KING (1989)).

Let $\{\lambda_j\}$ be the set of n eigenvalues of A, ordered in decreasing absolute value, then one may express Q as a linear combination of Chi-square random variables

$$Q = \sum_{j=1}^{n} \lambda_j \chi_j^2(1). \tag{2}$$

The distribution of Q can be found by numerically integrating the characteristic function of Q. However, when n is large, it is computationally demanding or even impossible to compute all eigenvalues. Even when they are known it can be costly to compute the exact distribution for large values of n. Therefore we approximate the distribution of Q in two ways.

Firstly, in section 2.1 we decompose Q into two terms

$$Q = Q_1 + Q_2 = \sum_{j=1}^{k} \lambda_j \chi_j^2(1) + \sum_{j=k+1}^{n} \lambda_j \chi_j^2(1) \tag{3}$$

where we retain the dominating λ's in Q_1 and replace the remaining part by a normal variable with the appropriate mean and variance. We write the resulting approximation as

$$Q^* = \sum_{j=1}^{k} \lambda_j \chi_j^2(1) + N(\mu, \sigma^2), \tag{4}$$

where for the expectation μ and the variance σ^2 of Q_2 one finds

$$\mu = \operatorname{tr} A - \sum_{j=1}^{k} \lambda_j, \quad \text{and} \quad \sigma^2 = 2\operatorname{tr} A^2 - 2\sum_{j=1}^{k} \lambda_j^2.$$

Secondly, in section 2.2 an iterative scheme adapted from SNEEK and DECROMBRUGGHE (1990) is given through which the dominating k eigenvalues of A in (1) are obtained (approximated) in only $\mathcal{O}(n^2)$ arithmetic operations. The values μ and σ^2 in (4) are obtained from $\operatorname{tr}(A)$ and $\operatorname{tr}(A^2)$ in $\mathcal{O}(n^2)$ operations also. If A is a sparse matrix (e.g. a band diagonal matrix), a low rank

matrix, or a combination of these types, then the number of required operations can sometimes be reduced to $\mathcal{O}(n)$.

In section 3 the accuracy of these approximations are shown for several different sets of $\{\lambda_j\}$ by comparing the approximate distribution with the exact one. Finally section 4 summarizes and gives some conclusions.

2. THEORY

2.1 Approximation to the distribution of a quadratic form

The distribution of quadratic forms in normal variables is (among many others) considered by KOOPMANS (1942), IMHOF (1961), KOERTS and ABRAHAMSE (1969), DAVIES (1980), PALM and SNEEK (1984) and FAREBROTHER (1984, 1990). Computations and algorithms are often based on the (numerical) inversion formula of GIL-PELAEZ (1951)

$$P(Q \leq c) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{t} \operatorname{Im}\left\{e^{-itc}\varphi(t)\right\} dt \tag{5}$$

where $\varphi(t)$ is the characteristic function of Q. Evaluating expression (5) has first received attention of IMHOF (1961) and later among others of KOERTS and ABRAHAMSE (1969), DAVIES (1973) and FAREBROTHER (1980, 1984, 1990).

Approximations to the distribution of Q are often based on some (asymptotic) expansion and they usually require knowledge about some low order moments of Q (see e.g. SNEEK (1983), ALI (1984) and the references made in these papers). If higher order moments are known then they can be used to find e.g. the Gram-Charlier, Edgeworth or Cornish-Fisher expansion or to fit e.g. the Pearson distribution to the first four moments (see KENDALL and STUART (1969) volume I and references therein).

If however the matrix A from expression (1) has a small number of eigenvalues that 'dominate' the set $\{\lambda_j\}$, then approximations are mostly not very successful. Especially when only the first two moments of Q are employed then e.g. skewness can not be captured. In this paper we assume that the dimensions of the matrix A are so large that computation of $\operatorname{tr} A^3$, needed for evaluating the third moment of Q is prohibitively expensive (as is any $\mathcal{O}(n^3)$ operation).

The characteristic function of Q in (2) equals

$$\varphi(t) = \prod_{j=1}^{n} (1 - 2i\lambda_j t)^{-\frac{1}{2}} \tag{6}$$

when the non-centrality parameters all equal zero. We will approximate $\varphi(t)$ by

retaining the first k terms in the product (k to be determined later) and by using a Taylor expansion around $t=0$ for the remaining terms:

$$\varphi(t) = \varphi_{Q_1}(t)\,\varphi_{Q_2}(t) = \prod_{j=1}^{k}(1-2i\lambda_j t)^{-\frac{1}{2}}\,\exp\{P(t)\}$$

where for convenience we write φ_{Q_2} in terms of an exponential function. It is well known that the form $\exp\{p(t)\}$, with $p(t)$ a polynomial, can only be a characteristic function if $p(t)$ is of at most second degree (MARCINKIEWICZ (1938)). Using a second order expansion in the exponent essentially means that Q_2 is approximated by a normally distributed variable. If the approximation Q^* is to capture skewness, curtosis and higher order deviations from normality, then this is done exclusively through Q_1, the part with the dominating eigenvalues. About the validity of the Taylor expansion we note that in practice one integrates in (5) only over some finite interval $0 \le u \le U$, so for small enough λ's the approximation will be good. Furthermore, the integrand mainly contributes for small values of u, the area where the Taylor expansion is most accurate. Finally, the smaller the value of an appropriate U, the larger the eigenvalues can be that are put into Q_2 without affecting the accuracy of the approximation too much; we come back to this point later.

Since one is integrating a proper characteristic function, the type of approximation that is proposed always leads to a proper cumulative density function (in contrast with the asymptotic expansions mentioned above).

In the next theorem we show that the approximation becomes arbitrarily accurate if $n \to \infty$ and $k = k(n) \to \infty$ at whatever slow rate. In practice the matrix A varies with n, so we should denote the eigenvalues of A as $\{\lambda_{j,n}\}_{j=1}^{n}$ to express the dependence on n explicitly. We suppress however this index n unless ambiguity arises.

Theorem Let $\{\lambda_j\}_{j=1}^{n}$ be an array of real numbers with $|\lambda_{j+1}| \ge |\lambda_j|$ for every $j = 1,\ldots,n$. If $n \to \infty$ and $k = k(n) \to \infty$ then Q^* in (4) converges in distribution to Q in (2).

proof We only outline the major points in the proof.

Choose $k(n)$ as a function of n and choose some (large) constant N. Let $\varepsilon > 0$ be given, then $\exists n_0$ such that $\forall n \ge n_0$ one has $k\varepsilon \ge N$. We distinguish two cases.

(i) $\sum_{k+1}^{n}\lambda_j^2 \le \varepsilon M$, with $M = \sum_{1}^{k}\lambda_j^2$.

In this case Q_1 dominates Q_2 and as Q_1 is handled exactly, the cumulative distribution function of Q^* is close to that of Q.

(ii) $\sum_{k+1}^{n}\lambda_j^2 > \epsilon M \geq \epsilon k\lambda_{k+1}^2 \geq N\lambda_{k+1}^2$.

In this case one can prove the condition of lemma 1 in the appendix.

Note that in the proof one does not use any characteristics of the sequence $\{\lambda_j\}_{j=1}^{\infty}$, so the approximation is uniform in the sequences $\{\lambda_j\}$. The implication is that all results carry over to double arrays: it is not possible to give a sequence of lambdas for each n such that the approximation does not improve if $n \to \infty$.

Also note that in the conditions of the theorem k grows as a function of n. Of course one may have that

$$\sum_{k+1}^{n}\lambda_j \chi_j^2(1) \longrightarrow N(\mu,\sigma^2)$$

for a finite (fixed) value of k if $n \to \infty$, especially if $\lambda_j = \lambda_{j,n}$. Some conditions for convergence of Q_2 to normality are given in the appendix and this is sufficient to ensure convergence of Q^* to Q. In practice one hopes that moderate values for k suffice to give reasonable approximations to the exact distribution of the quadratic forms. There exist cases where both Q_1 will dominate Q_2 and Q_2 converges to a normally distributed variable if $k \to \infty$; under these circumstances one would expect the approximation to work particularly well.

In practice one needs a criterion to determine when to stop increasing the value of k. It is difficult to relate the value of k directly with the error that is made when approximating the distribution of Q as accuracy strongly depends on the particular sequence $\{\lambda_j\}$. Assuming that the sequence has been normalized such that $\Sigma_1^n\lambda_j^2 = 1$, it seems reasonable that terms with small $|\lambda_j|$ should be put into Q_2. On the other hand, the value of $|\lambda_j|$ seems less important if Q_2 is dominated anyway by Q_1. We have taken the following approach: increase the value of k until

$$\lambda_k^2(1-\Sigma_1^k\lambda_j^2) \leq \delta\ \Pi_1^k(1+2.5\lambda_j^2) \qquad (7)$$

for some $\delta > 0$. Clearly k is no longer increased if $|\lambda_k|$ is small, but other factors are important as well. A factor $(1-\Sigma_1^k\lambda_j^2)/a$ indicates how many more nonzero λ's are to come if they are all equal to a (the most unfavourable case as far as the cumulants of Q_2 are concerned); $(1-\Sigma_1^k\lambda_j^2)$ 'weakens' the condition of putting small λ's into Q_2 if Q_1 is dominating Q_2. The factor $\Pi_1^k(1+2.5\lambda_j^2)$ is based on the empirical fact that U, the upperbound of the integration interval used in (5), tends to decrease if the number of λ's increases, but the validity of the replacement of Q_2 by a normal variable is related to the

validity of the Taylor expansion of the characteristic function of Q_2 in the range $0 \le u \le U$. This implies that the smaller the value of U, the larger the values of the λ's can be without distorting the approximation too much. The value 2.5 was obtained after some numerical experimentation. In section 3 we provide evidence that the value of δ can be used as a measure of accuracy of the approximation. This is important for practice, even if the value of k in (7) is predetermined, because in the latter case one may compute a (maximum) value for δ such that the fixed value of k would have been found.

We finally review in this section how to compute the distribution of Q^*. Collecting terms with equal eigenvalues to make the formulas compatible with e.g. those of IMHOF (1961), denote the set of m distinct eigenvalues of A in (1) as $\{\lambda_j\}$ with orders of multiplicity $\{m_j\}$ and rewrite Q as

$$Q = \sum_{j=1}^{m} \lambda_j \chi^2(m_j) .$$

Very similar to IMHOF (1961) one finds

$$F(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{u \, \rho(u)} \, du \tag{8}$$

where now however

$$\theta(u) = \frac{1}{2} \sum_{j=1}^{m} \left[m_j \operatorname{arctg}(\lambda_j u) + \frac{\delta_j^2 \lambda_j u}{1+\lambda_j^2 u^2} \right] - \frac{1}{2} x u + \frac{1}{2} \mu u$$

$$\rho(u) = \prod_{j=1}^{m} (1+\lambda_j^2 u^2)^{\frac{1}{4} m_j} \exp\left[\frac{1}{2} \sum_{j=1}^{m} \frac{\delta_j^2 \lambda_j^2 u^2}{1+\lambda_j^2 u^2} \right] \exp\left[\frac{1}{8} \sigma^2 u^2 \right].$$

We note that in DAVIES (1973) an algorithm is given to compute $F(x)$ from (8) by numerical integration, which is carried out over a finite range $0 \le u \le U$. Since there is a normal part in in the integrand one usually has a much sharper bound for the truncation error

$$t_U = \frac{1}{\pi} \int_U^\infty \frac{\sin \theta(u)}{u \, \rho(u)} \, du \tag{9}$$

than the one obtained by KOERTS and ABRAHAMSE (1969). The derivation is as follows:

$$\pi \, |t_U| \le \int_U^\infty \frac{1}{u} \prod_{j=1}^{k} (1+\lambda_j^2 u^2)^{-\frac{1}{4} m_j} \exp(-\tfrac{1}{8} \sigma^2 u^2) \, du \tag{10}$$

$$\le \prod_{j=1}^{k} (1+\lambda_j^2 U^2)^{-\frac{1}{4} m_j} \frac{1}{U} \int_U^\infty \exp(-\tfrac{1}{8} \sigma^2 u^2) \, du = \prod_{j=1}^{k} (1+\lambda_j^2 U^2)^{-\frac{1}{4} m_j} \frac{2}{U\sigma} \left[1 - \Phi(\tfrac{1}{2} \sigma U) \right] \tag{11}$$

where Φ is the cumulative normal distribution function.
By taking $\exp(-\frac{1}{8}\sigma^2u^2)$ out of the integrant in (10) one more or less arrives at the bound of KOERTS and ABRAHAMSE (1969). DAVIES (1980, formula (8) with his U equal to $\frac{1}{2}U$ in our formula) replaces $\frac{u}{U}$ by 1 for $u \geq U$ in (9) and gives the following bound

$$\prod_{j=1}^{k}(1+\lambda_j^2U^2)^{-\frac{1}{4}m_j}\frac{1}{U^2}\int_U^\infty u\exp(-\tfrac{1}{8}\sigma^2u^2)du = \prod_{j=1}^{k}(1+\lambda_j^2U^2)^{-\frac{1}{4}m_j}\frac{4}{U^2\sigma^2}\exp(-\tfrac{1}{8}\sigma^2U^2)$$

but this is always larger than (11). In (9) the term

$$C(u) = \exp\left[-\frac{1}{2}\sum_{j=1}^{k}\frac{\delta_j^2\lambda_ju^2}{1+\lambda_j^2u^2}\right]$$

is left out because $C(u)$ is monotonically decreasing.

2.2 A generalized power method to compute eigenvalues

In order to get a feasible approximation in practice for the distribution of Q in (1) we need a computationally cheap way to find the dominating eigenvalues of A as we envisage applications in which the order of the matrix is large. In PALM and SNEEK (1984) and FAREBROTHER (1985, 1990) an eigenvalue-free method is suggested, but that method still requires tridiagonalization of A, an $\mathcal{O}(n^3)$ process in general.
In approximating the exact distribution of the quadratic form one may wonder what the effect is of using approximations to the dominating lambdas in Q_1. If λ_j is estimated as $\tilde{\lambda}_j$, then the characteristic function is perturbed by the term (denoting all eigenvalues as single again)

$$\left[\frac{1-2i\tilde{\lambda}_jt}{1-2i\lambda_jt}\right]^{-\frac{1}{2}} = \exp\left[-\frac{1}{2}\left\{(\lambda_j-\tilde{\lambda}_j)\frac{(2it)}{1} + (\lambda_j^2-\tilde{\lambda}_j^2)\frac{(2it)^2}{2} + (\lambda_j^3-\tilde{\lambda}_j^3)\frac{(2it)^3}{3}+\ldots\right\}\right] \quad (12)$$

The first two terms in the exponent are corrected by the normal part in Q^*, so

$$\exp\left[-\frac{1}{2}\sum_{j=1}^{n}(\lambda_j^3-\tilde{\lambda}_j^3)\frac{(2it)^3}{3}\right]$$

is the first term that remains if eigenvalues $\{\lambda_j\}$ *in* (4) are approximated by $\{\tilde{\lambda}_j\}$. From $(\lambda_j^3-\tilde{\lambda}_j^3)=(\lambda_j-\tilde{\lambda}_j)(\lambda_j^2+\lambda_j\tilde{\lambda}_j+\tilde{\lambda}_j^2)$ it follows that the eigenvalues with the largest absolute value should be known most accurately. Note that for small eigenvalues we take $\tilde{\lambda}_j=0$, though mean and variance are corrected.

In SNEEK and DECROMBRUGGHE (1990) an acceleration method for iterative model solution is developed and as one of the byproducts they obtain the dominating

eigenvalues of a matrix A if the system $Ay=b$ is solved for y. In their context the occurrence of multiple eigenvalues is not relevant and their method only delivers single eigenvalues. We adapt there method by introducing a sort of deflation on A using the eigenvalues and eigenvectors that already have been found. In order to keep this paper self contained we review some of the theory.

Denote the set of eigenvalues of the symmetric matrix A ordered in decreasing absolute value by $\{\lambda_j\}$, $j=1,...,n$. Suppose that A has n linearly independent eigenvectors $\{e_j\}$ and let the eigenvectors be chosen such that they form an orthonormal system. Let the random vector $y^{(0)}$ of unit length be given and generate the sequence of iterates $\{y^{(0)},y^{(1)},y^{(2)},...\}$, where $y^{(t+1)}=Ay^{(t)}$. To warm up the process it may be necessary to discard the beginning of the sequence, in which case for convenience we renumber the sequence.

The vectors $y^{(0)}$ and $y^{(t)}$ can be written as

$$y^{(0)}=\sum_{j=1}^{n}\beta_j e_j, \qquad\qquad y^{(t)}=\sum_{j=1}^{n}\lambda_j^t\beta_j e_j.$$

Suppose that some eigenvalues are dominant, say the first k, then one can write

$$y^{(t)}=\sum_{j=1}^{n}\lambda_j^t\beta_j e_j \;\approx\; \sum_{j=1}^{k}\lambda_j^t\beta_j e_j \;, \tag{13}$$

where k is much smaller than n, in practice always less than 10 if warming up has taken place long enough. This implies that the sequence $\{y^{(t)}\}$ approximately satisfies a k-th order homogeneous (vector) difference equation, i.e. there exists a vector $\varphi=(\varphi_1,\varphi_2,...,\varphi_k)'$ such that

$$y^{(t)} - \varphi_1 y^{(t-1)} - ... - \varphi_k y^{(t-k)} \approx 0. \tag{14}$$

We note that there is a 1-1 correspondence between the values $\lambda_1,...,\lambda_k$ in (13) and the coefficients $\varphi_1,...,\varphi_k$. The coefficient vector φ, specializing to the case of a symmetric matrix A, can be computed as follows.

Let the sequence $\{s_t\}_{t\geq 1}$ be defined as

$$s_{2t-1}=(y^{(t-1)})'y^{(t)}, \qquad s_{2t}=(y^{(t)})'y^{(t)}.$$

From (13) it is evident that the scalar sequence $\{s_t\}$ obeys the same difference equation as the sequence $\{y^{(t)}\}$ does. Note that there are several alternatives to construct sequences $\{s_t\}$ with similar properties. For instance, one may take just one component of $y^{(t)}$ or take any linear combination of the components of $y^{(t)}$. As noted by SNEEK and DECROMBRUGGHE (1990) however these choices may lead to a lower order difference equation for $\{s_t\}$.

Define the Hankel matrix H and the matrix $S_{k,i}$ by

$$H = \begin{bmatrix} s_1 & s_2 & s_3 & \cdots \\ s_2 & s_3 & s_4 & \cdots \\ s_3 & s_4 & s_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad \text{and} \quad S_{k,i} = \begin{bmatrix} s_{1+i} & s_{2+i} & \cdots & s_{k+i} \\ s_{2+i} & s_{3+i} & \cdots & \\ \vdots & \vdots & \ddots & \\ \vdots & \vdots & \ddots & \\ s_{k+i} & s_{k+1+i} & & s_{2k-1+i} \end{bmatrix}$$

The matrices $S_{k,i}$ are finite sub matrices of H. Let $det_{k,i} = \det(S_{k,i})$. If the sequence $\{s_t\}$ satisfies a m-th order difference equation but not a $(m-1)$-th order difference equation, then $det_{k,i} = 0$ for $k \geq m+1$, but $det_{m,i} \neq 0$. In practice the difference equations are not holding exactly and furthermore the order of an approximate difference equation will be lower as we progress through the sequence $\{s_t\}$, because in (13) the large lambdas become more and more dominant. In BEGUIN, GOURIEROUX and MONTFORT (1980) a detailed discussion (in the context of determining the order of ARMA(p,q) models) is given of the corner method, which can be used to determine the order m and the point in the sequence where the difference equation starts to hold. Knowing the (approximate) order m and possibly after renumbering the sequence $\{s_t\}$ the vector φ can be calculated from

$$\begin{pmatrix} \varphi_m \\ \varphi_{m-1} \\ \vdots \\ \varphi_1 \end{pmatrix} = \begin{bmatrix} s_1 & s_2 & \cdots & s_m \\ s_2 & s_3 & \cdots & \\ \vdots & \vdots & \ddots & \\ s_m & s_{m+1} & & s_{2m-1} \end{bmatrix}^{-1} \begin{pmatrix} s_{m+1} \\ s_{m+2} \\ \vdots \\ s_{2m} \end{pmatrix}.$$

The λ's can finally be calculated from φ.

As an alternative to this computation SNEEK and DECROMBRUGGHE (1990) note that if $s_i = \Sigma_1^m b_k \lambda_k^i$ then one has for any $\mu \in \mathbb{R}$

$$s_i - \mu s_{i-1} = \sum_{k=1}^{m} b_k \lambda_k^{i-1} (\lambda_k - \mu),$$

so the sequence $\{s_i - \mu s_{i-1}\}$ satisfies the same difference equation if μ is different from all lambdas, but it satisfies a difference equation of order $m-1$ if μ equals λ_k for some $k = 1,...,m$. This leads to the conclusion that solution of the generalized eigenvalue problem

$$|S_{m,1} - \mu S_{m,0}| = 0 \qquad (15)$$

exactly entails the values $\lambda_1, \lambda_2, ..., \lambda_m$. As an alternative to (15) one may compute the eigenvalues of the matrix $S_{m,0}^{-1} S_{m,1}$.

One problem not mentioned in SNEEK and DECROMBRUGGHE (1990) is that the procedure above will only lead to distinct eigenvalues $\{\lambda_1,...,\lambda_m\}$. This

follows from the representation of $y^{(t)}$. If e.g. $\lambda_1=\lambda_2$, then the eigenvectors $\{e_1,e_2\}$, corresponding to λ_1 and λ_2, can be recombined to the eigenvectors $\{e_1^*,e_2^*\}$ such that $y^{(0)}$ has no component in the direction of e_2^*:

$$\lambda_1^t\beta_1 e_1+\lambda_2^t\beta_2 e_2 \;=\; \lambda_1^t(\beta_1 e_1+\beta_2 e_2) \;=\; \lambda_1^t\beta_1^* e_1^*.$$

To find λ_2 in this case one may iterate with the matrix $A-\lambda_1 e_1 e_1'$ or iterate with the matrix A and set the iterates $y^{(t)}$ through Gram-Schmidt orthogonalization perpendicular to e_1.

To find the eigenvectors $e_1,\ldots,e_m$ we observe that

$$Y \;\approx\; [\Sigma\lambda_j e_j \,|\, \Sigma\lambda_j^2 e_j \,|\, \ldots \,|\, \Sigma\lambda_j^m e_j] \;=\; [e_1 \,|\, e_2 \,|\, \ldots \,|\, e_m]\,\Lambda \qquad (16)$$

where

$$Y \;=\; [y^{(1)} \,|\, y^{(2)} \,|\, \ldots \,|\, y^{(m)}] \qquad \text{and} \qquad \Lambda \;=\; \begin{bmatrix} \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^m \\ \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^m \\ \vdots & \vdots & \ddots & \\ \lambda_m & \lambda_m^2 & & \lambda_m^m \end{bmatrix}.$$

Solving (16) leads to

$$E \;=\; Y\Lambda^{-1}$$

In practice we keep all but the last column of E as a set of eigenvectors and restart the iteration process perpendicular to the already obtained eigenvectors. The starting value for $y^{(0)}$ is taken as a linear combination of the discarded eigenvector and a spherically symmetric distributed random vector. One may have to repeat this procedure if higher multiplicities are present in the set of eigenvalues, though if multiplicities are very high then the normal approximation would probably not be so bad and thus the eigenvalues would not be needed at all.

The method described so far can be considered as a sort of generalized power method and usually the approximations deteriorate quickly as the eigenspace used for deflation is not approximated well enough. To get some idea about the orders of magnitude involved consider the iterations $Ax^{(i)} = x^{(i+1)}$ and let the orthogonal matrix U be such that

$$U'AU \;=\; B \;=\; \text{diag}(\lambda_1,\ldots,\lambda_n).$$

Writing $y^{(i)}=U'x^{(i)}$ and premultiplying the equations by U' one may consider the equivalent iterations

$$By^{(i)} \;=\; y^{(i+1)}.$$

(If a random vector is generated anywhere we use spherically symmetric

distributions, so orthogonal transformations do not change the principles of the procedure). Let

$$y^{(t)} = \sum_{i=1}^{n} \alpha_i \lambda_i^t u_i = \sum_{i=1}^{m} \alpha_i \lambda_i^t u_i + r^{(t)}$$

where $\{u_i\}$, $i=1,...,n$, are the eigenvectors of B. It follows that

$$\left[u_1 \ \dots \ u_m\right] = \left[y^{(1)} \ \dots \ y^{(m)}\right] \Lambda^{-1} - \left[r^{(1)} \ \dots \ r^{(m)}\right] \Lambda^{-1}. \qquad (17)$$

The vector $r^{(t)}$ can be written as $(0 \dots 0 \ \alpha_{m+1}\lambda_{m+1}^t \dots \alpha_n \lambda_n^t)'$ in which the first m elements are zero. It is the last part of (17) that is the error in the estimate of the eigenspace spanned by $\{u_1,...,u_m\}$. The coefficients however involve only terms in powers of $\lambda_{m+1},\lambda_{m+2},...,\lambda_n$. Even though the lambdas in (17) are unknown, the space spanned by $y^{(1)},...,y^{(m)}$ is the same as the space spanned by the columns of $\left[y^{(1)}...y^{(m)}\right] \Lambda^{-1}$. Although the approximate eigenvectors and the exact eigenvectors may differ substantially, what matters is that the spanned spaces are reasonably close to one another.
In practice we found that if only a few eigenvalues were quite dominating, then we got more accuracy than in the case where many eigenvalues were approximately equally large. Fortunately, in the former case we indeed need to know Q_1 in (3) more accurately than in the latter, where the normal part may take over.

The considerations above do not give a clue about the accuracy in the determination of the lambdas themselves. In that case one would have to consider the eigenvalue problem in (15) where the $\{s_i\}$ are determined from the iterates $\{y^{(i)}\}$. Although it is possible to make some order of magnitude estimations about the vector φ, we were not able to do the same about the eigenvalues. In practice the accuracy seems well enough for our purposes as we try to show in section 3.

3. Empirical Results

In this section we show how the approximations work in practice. To do so we consider several types of patterns for the eigenvalues $\{\lambda_i\}$ and investigate the approximations for n=40, 80 and 160. We do not report the results for all values of n as sometimes the results hardly vary in n.
The typical sets of $\{\lambda_i\}$ we consider are the folowing:

set (1) $\lambda_i = 1 - \frac{i}{n+1}$, $i=1,\ldots,n$

set (2) $\lambda_i = -\log\left(\frac{i}{n+1}\right)$, $i=1,\ldots,n$

set (3) $\lambda_i = \log\left(1-\frac{i}{n+1}\right) + \log(n+2)$, $i=1,\ldots,n$

set (4) $\lambda_i = 0.5 - \frac{i}{n+1}$, $i=1,\ldots,n$

set (5) $\lambda_i = \pm\log\left(\frac{i}{n/2+1}\right)$, $i=1,\ldots,n/2$

set (6) $\lambda_i = \pm\log\left(1-\frac{i}{n/2+1}\right) \mp \log(n/2+2)$, $i=1,\ldots,n/2$

set (7) $\lambda_i = 1.1^{-i}$, $i=1,\ldots,n$

set (8) $\lambda_i = 1.05^{-i}$, $i=1,\ldots,n$

set (9) $\lambda_i = i^{-1}$, $i=1,\ldots,n$

set (10) $\lambda_i = i^{-0.75}$, $i=1,\ldots,n$.

If one views the sets as percentile points from a density, then set (1) corresponds to a uniform distribution, set (2) to a negative exponential distribution, set (3) to the mirror image of (2) but shifted to the right. Sets (4)-(6) are like (1)-(3) but including their mirror image for negative λ's, which makes them symmetric around zero.

For the sets (1)-(6) the central limit theorem applies if $n \to \infty$ and this implies that given k the use of k dominant eigenvalues becomes less influential if $n \to \infty$. The sets (1)-(3) above were also shifted to the left and right by adding a constant to all the λ's, but as in all cases the approximations improved or remained very similar we do not report the results.

In table 1 we report results when Q in (3) is approximated by Q^* in (4). The table is organized as follows.

The columns 2-11 correspond to the different sets $\{\lambda_i\}$ and given n and δ and given the value p_{target} ('target' significance level) listed in the first column they are ontained by solving x and p_{actual} from

$$\left.\begin{aligned} P(Q^* \le x) &= p_{target} \\ P(Q \le x) &= p_{actual} \end{aligned}\right\} \qquad (18)$$

Note that this corresponds to hypothesis testing in practice when one uses the approximation but would be interested to know the actual significance level. Finally we compute and report $(p_{actual} - p_{target})$ multiplied by 10000 to facilitate comparison of the results. For each set $\{\lambda_j\}$ we also report the

value k, the number of eigenvalues needed in the approximation for the given values of n and δ. For the normal approximation (i.e. $\delta=0$) k obviously always equals zero. For example, in table 1A for set (1) and $n=40$ the (normal) approximation would indicate a left tail probability 0.005, but the true significance level in that case equals 0.0001, a difference of -49 times 10^{-4}. Note that an entry with ± 100 means that the true level is missed by a full percent.

In table 1 we selected the results for the normal approximation ($\delta=0$), for $\delta=0.006$ and k set at a maximum of 12 and for $\delta=0.001$. The value $\delta=0.006$ is chosen because the approximations are reasonable enough to be valuable in practice; the limit in k was set at 12 to keep the approximation within a reasonable computational burden. We present some results for $\delta=0.001$ to show on the one hand that the approximation indeed improves, but on the other that k may increase considerably (even beyond practical limits if the eigenvalues are unknown) in cases where the central limit theorem is taking over. We note that the user may set the maximum value for k and compute the (largest) δ that would be required to get that k. A very small value for δ would indicate that the approximation is very accurate (see also table 2).

From tabel 1A it is evident that the normal approximation can be grossly inadequate for the sets (1)-(3) and (7)-(10) even for n=160, though in sets (1)-(3) one can see the central limit theorem already at work. From tabel 1B and 1C it can be seen that using Q^* considerably improves upon the normal approximation. Especially for the sets (7)-(10) an approximation based on $\delta=0.001$ is practically feasible because $k\leq 15$ and gives acceptable results except perhaps for $p_{target}\leq 0.01$. For sets (1)-(3) a value $\delta=0.006$ is near the smallest that is feasible, though we found already a case $k=17$. As the large k-values always occurred in sets(1)-(6) and resulted in rather small errors we decided to show table 1A where k was restricted to 12. The approximation is acceptable for $p_{target}\leq 0.05$. It is tempting to conclude that the approximation improves if n gets larger, but that information is useless for practice as one always can increase n without essentially changing the characteristics of the sequence $\{\lambda_j\}$ by adding zeros.

For sets with positive λ's the approximation is somewhat less accurate for the p-values 0.005, 0.01 and 0.025. This corresponds to percentiles close to zero for which the integrand (8) oscillates slowly (perhaps the upperbound U is quite large in these cases, so the Taylor expansion only works well for terms with very small λ's). If one knows that one is dealing with a positive (semi) definite quadratic form, then fortunately only the right hand side tail

of the distribution is usually needed. In this paper however we do not consider positive definite forms as a special case.
For the sets (7)-(10) the results hardly vary in n, which is to be expected since very quickly the terms become negligible. Also quite obvious is the fact that the normal approximation improves considerably if n gets larger, especially for sets (4)-(6), but markedly as well for sets (1)-(3).

To indicate the relation between the value of δ in (7) and the (maximum) error that is made using the distribution of Q^* instead of Q we have constructed table 2. To compile the information in table 2 we compute, given the value of n, for each set $\{\lambda_i\}$ the maximum of $|p_{target} - p_{actual}|$ for all p-values that are considered (see first columns in table 1). For the case at which the maximum occurres we report the maximum e_{max}, the value p_{target} and the value of k that is used in the approximation. Table 2 is obtained by maximizing over all sets $\{\lambda_i\}$ that are considered.
In table 2A the first column contains the value of δ and given that value the table should be read from left to right. Column (2) indicates at what value of p_{target} the maximum occurs (this hardly varies with n), columns (3) shows the maximum errors obtained from all sets and all p-values but given the value of n. Columns (4) and (5) show the corresponding set $\{\lambda_j\}$ and the number of eigenvalues k used in the approximation for the case where the maximum is obtained. For n=80 and n=160 there are similar columns. In table 2B we have reported similar figures but with the value of k fixed.
From the table one sees that the maximum error does not vary much with n and for small values of δ always set (9) leads to the maximum error. Also for the reported cases the maximum is always obtained for the positive definite sets and in the left tail of the distribution ($p_{target} \leq 0.05$). Not visible from table 2, though likely from table 1C ($\delta = 0.001$), is the fact that for sets (1)-(6) the value of k becomes prohibitively large if δ is less than say 0.006; as the maximum errors for values of δ in this range are obtained in sets (7)-(10) one may conclude that our criterion to increase the value of k in (7) could be changed somewhat towards disfavouring large values of k. However, even then feasible values for (very) small δ's are only obtained for sets (7)-(10).
Note in table 2A and 2B the important special cases $\delta = 0.006$ and $k = 12$ at which the maximum deviation between target and actual significance level is approximately 1 percent. From these figures we would in practice advice to use the value $k = 12$ and compute the corresponding δ as a measure for the accuracy of

the approximation, or if the computer budget is really tight to use the combination $k = 12$ and $\delta = 0.006$.

For the results sofar we used the exact eigenvalues and did not bother about how to find those eigenvalues. Usually the bulk of the work of computing tail probabilities of quadratic forms is in finding the eigenvalues. We approximated the eigenvalues as explained in section 2.2. For the practical implementation we have made the following choices.
One has to decide when the determinants $det_{k,i}$ are effectively zero. First the sequence $\{s_t\}$ is normalized such that $s_1 = 1$ and then the matrix A (and consequently each eigenvalue) is multiplied by a factor such that $s_3 = (y^{(2)})'y^{(2)} = 1$; this is to avoid badly scaled hankel matrices H. In addition we also scale each $S_{k,i}$ individually for each k by making each of its columns a vector of unit length times $(\pi/2)$; this factor ensures that if the columns of $S_{k,i}$ point in (sperically symmetric) random directions, then the expected value of $det_{k,i}$ will not decrease with k. The matrix $S_{k,i}$ is updated as soon as a new iterate $y^{(t)}$ is available and each time the determinants of all lower right corners of $S_{k,i}$ are computed. The determinant $det_{m+1,i}$ is taken to be zero (an m-th order difference equation is acceptable) if a quotient $det_{m+1,i-1}/det_{m,i}$ is less than 10^{-8}; if more quotients meet this criterion then the smallest quotient determines the value of m. The problem of multiple eigenvalues is handled by restarting the iteration process two more times as explained in section 2.2.

In table 3 we report as typical examples the approximation of the eigenvalues of the sets (2), (3) and (4) and the total number of iterates $y^{(t)} = Ay^{(t-1)}$ that were required to calculate these approximations. For other sets eigenvalues and other values of n the results are very similar and are not reported here. Clearly the strongly dominating eigenvalues (set (2)) are most accurate. If one goes down the list of eigenvalues however then the approximation becomes poorer. Partly the reason for this is that the procedure has 'skipped' some eigenvalues from the list and this problem can be alleviated to some extend by restarting the iteration process more often. We did not do so since we wanted to keep the number of iterations small and because the approximations are good enough for our purposes. In our experience somewhere between 30 and 40 iterations is usually enough to obtain 12 eigenvalues and usually between 20 and 25 iterations suffice to obtain 4 eigenvalues.
We note that we did not shift the eigenvalues of A to the right or left to get

better dominance properties or use Chebyshev polynomials during the warming up phase to remove non-dominant factors faster. Perhaps these techniques may slightly improve the procedure. If much greater accuracy is required for the determination of the eigenvalues, then perhaps Bauer's simultaneous iteration method can be a solution, though usually the computational burden will increase considerably.

In table 4 we show some selected results for the case taht Q is replaced by Q^* and for the combined approximation procedure, i.e. the dominating eigenvalues λ_j are approximated by $\tilde{\lambda}_j$ and part of the quadratic form is replaced by a normally $N\left(\mathrm{tr}A-\Sigma\tilde{\lambda}_j, 2\mathrm{tr}A^2-2\Sigma\tilde{\lambda}_j^2\right)$ distributed variable.
In this table we fix the value of k rather than the value of δ as in table 1. The first column again contains the target significance level and the remaining columns contain the actual signicifance level if the approximation is used in practice (i.e. we solve (18)). Note that in the second column $k=0$ corresponds to the normal approximation and the information in this column is contained in table 1A as well. In the columns 3 and 4 we have used k=4 dominant eigenvalues in the approximation; in column 3 the exact values are used and in column 4 the approximated eigenvalues $\tilde{\lambda}_j$ are used. The remaining columns are organized similarly.
The most important fact revealed from table 4 (and other not reported results) is that the use of approximated λ's does not change the approximation as a whole very much if $k \leq 12$. This is actually the reason why we show the full results only for the sets (1) and (2) as essentially the information in tables 4A-4C is similar to that in table 1B for $\delta = 0.006$. As for the sets (7)-(10) the results are virtually the same for all n we show only $n=80$ in table 4C.

We checked similar results corresponding to $k=16$ and $k=20$, but then the approximation does improve if the exact eigenvalues are used but remains very similar if approximated values are used. The conclusion obviously is that if more accuracy is needed in approximating the tail probabilities, then the procedure of finding the eigenvalues must be enhanced. In another paper we deal with this problem using a combination of our method and the simultaneous iteration method from Bauer (see RUTISHAUSER (1969)) for a case when only $O(n)$ arithmetic operations are needed to calculate the Hankel matrices $S_{k,i}$.

4. CONCLUSIONS

In this paper we have given evidence that the computation of the distribution of large quadratic forms in normal variables is feasible in practice to within 1 percent accuracy. This seems good enough if e.g. one is interested in testing at the 5 percent significance level. Furthermore, the procedure enables the user to monitor the accuracy to some extend using the quantity δ from (7), so in practice it is possible in certain cases to use the method if testing is required at a lower significance level; this is especially true if the quadratic form is dominated by relatively few eigenvalues and its distribution is far from normal.

The approximation will usually require less than $40n^2$ multiplications to obtain approximate eigenvalues and a numerical integration that does not depend on n. In special cases the number of arithmetic operations may drop down to $\mathcal{O}(n)$ if the quadratic form has special structure. In such cases it may be worthwhile to increase the accuracy in the approximation of the eigenvalues, but this needs further research. Also increasing the accuracy in special cases if e.g. third or higher order moments are known needs further research.

6. REFERENCES

ALI, M.M. (1984), "Distributions of the Sample Autocorrelations when Observations are from a Stationary Autoregressive-Moving-Average Process", *Journal of Business and Economic Statistics*, **2**, pp 271-278.

BEGUIN, J.-M., C. GOURIEROUX, and A. MONFORT (1980), "Identification of a mixed autoregressive-moving average process: the corner method", in *Time Series*, O.D. Anderson, ed., North-Holland Publishing Company.

DAVIES, R.B. (1973), "Numerical inversion of a characteristic function", *Biometrika*, **60**, 415-417.

DAVIES, R.B. (1980), "The distribution of a linear combination of χ^2 random variables", *Applied Statistics*, **29**, 323-333.

DUFOUR, J-M. and ROY, R. (1985), "Some Robust Exact Results on Sample Autocorrelations and Tests of Randomness", *Journal of Econometrics*, **29**, pp 257-273.

EVANS, G.B.A. and N.E. SAVIN (1981), "Testing for Unit Roots 1", *Econometrica*, **49**, pp 753-779.

EVANS, G.B.A. and N.E. SAVIN (1984), "Testing for Unit Roots 2",

Econometrica, **52**, pp 1241-1269.

FAREBROTHER, R.W. (1980), "The Durbin-Watson test for serial correlation when there is no intercept in the regression", *Econometrica*, **48**, 1553 – 1564.

FAREBROTHER, R.W. (1984), "The distribution of a positive linear combination of χ^2 random variables", *Applied Statistics*, **33**, 332-339.

FAREBROTHER, R.W. (1985), "Eigenvalue-free methods for computing the distribution of a quadratic form in normal variables", *Statistische Hefte*, **26**, 287-302.

FAREBROTHER, R.W. (1990), "The distribution of a quadratic form in normal variables", *Applied Statistics*, **39**, 294-309.

GIL-PELAEZ, J. (1951), "Note on the inversion theorem", *Biometrika* **38**, 481-2.

IMHOF, J.P. (1961), "Computing the distribution of quadratic forms in normal variables", *Biometrika*, **48**, 419-426.

KENDALL, M.G., and A. STUART (1969), *The advanced theory of statistics*, vol. 1, Charles Griffin & Company Limited, London.

KIEVIT, J.F. (1980), "Effects of ARMA Errors on Tests for Regression Coefficients: Comments on Vinod's Article; Improved and Additional Results", *JASA*, **75**, pp 353-358.

KING, M.L. (1989), "Testing for Fourth-order Autocorrelation in Regression Disturbances when First-order Autocorrelation is Present", *Journal of Econometrics*, **41**, pp 285-301.

KOERTS, J., and A.P.J. ABRAHAMSE (1969), *On the theory and application of the general linear model*, Rotterdam University Press, Rotterdam.

KOOPMANS, T (1942), Serial correlation and quadratic forms in normal variables, *Annals of Mathematical Statistics*, **13**, 14-33.

MARCINKIEWICZ, J. (1938), Sur une propriété de la loi Gauss, *Math. Zeitschr.*, **4**, 612.

PALM, F.C. and J.M. SNEEK (1984), "Significance tests and spurious correlation in regression models with autocorrelated errors", *Statistische Hefte*, **25**, 87-105.

RUTISHAUSER, H. (1969), "Computational Aspects of F.L. Bauer's Simultaneous Iteration Method", *Numer. Math.*, 13, 4-13.

SNEEK, J.M. (1983), "Some approximations to the exact distribution of sample autocorrelations for autoregressive moving average models", in *Time Series Analysis: Theory and Practice* 3, O.D. Anderson, ed., North-Holland Publishing Company, 265-289.

SNEEK, J.M., and D. DECROMBRUGGHE (1990), "Iterative model solution and automatic multi parameter convergence acceleration", *manuscript*.

Appendix

lemma 1 If $(\lambda_{k+1,n}^2/\Sigma\lambda_{j,n}^2) \longrightarrow 0$ then Q_2 converges to a normaly distributed variable.

proof

The cumulants κ_r of $Q_2 = \sum_{j=k+1}^{n} \lambda_{j,n}\chi_j^2(1)$ are equal to

$$\kappa_r = 2^{r-1}(r-1)! \sum_{k+1}^{n} \lambda_{j,n}^r.$$

From this the cumulants κ_r' of the standardized variable are for $r \geq 3$

$$\kappa_r' = \frac{2^{r-1}(r-1)!\ \Sigma\ \lambda_{j,n}^r}{\left[2\Sigma\lambda_{j,n}^2\right]^{\frac{1}{2}r}}.$$

The lemma now follows from the following inequalities for $r \geq 3$:

$$\frac{\sum_{k+1}^{n} \lambda_{j,n}^r}{\left[\sum_{k+1}^{n} \lambda_{j,n}^2\right]^{\frac{1}{2}r}} = \frac{\Sigma(\lambda_{j,n}/\lambda_{k+1,n})^r}{\left[\Sigma(\lambda_{j,n}/\lambda_{k+1,n})^2\right]^{\frac{1}{2}r}} \leq \frac{\Sigma(\lambda_{j,n}/\lambda_{k+1,n})^2}{\left[\Sigma(\lambda_{j,n}/\lambda_{k+1,n})^2\right]^{\frac{1}{2}r}} =$$

$$= \frac{\lambda_{k+1,n}^{-2}\ \Sigma\lambda_{j,n}^2}{\lambda_{k+1,n}^{-r}\left[\Sigma\lambda_{j,n}^2\right]^{\frac{1}{2}r}} = \lambda_{k+1,n}^{r-2}/\left[\Sigma\lambda_{j,n}^2\right]^{\frac{1}{2}r-1} \leq \left[\lambda_{k+1,n}^2/\Sigma\lambda_{j,n}^2\right]^{\frac{1}{2}r-1}$$

lemma 2 If either of the following conditions holds if $n \longrightarrow \infty$:

(i) $\sum_{j=1}^{k} \lambda_{j,n}^2 / \sum_{j=1}^{n} \lambda_{j,n}^2 \longrightarrow 1$

(ii) $Q_2 = \sum_{j=k+1}^{n} \lambda_{j,n}\chi_j^2(1) \longrightarrow N(\mu,\sigma^2)$

then the approximation becomes arbitrary accurate:

proof If the first condition holds then Q_2 becomes negligible compared to Q_1 as one can prove that $P(|Q_2| \leq \varepsilon|Q_1|) \rightarrow 1$. for all $\varepsilon > 0$. If the second condition holds then there is nothing to prove.

lemma 3 Let the sequencess $\{\lambda_{j,n}\}_{j=1}^{n}$ be normalized such that $|\lambda_{1,n}| = 1$. If $\sum \lambda_{j,n}^2 \longrightarrow \infty$, $n \longrightarrow \infty$ then for *any* bounded sequence $k(n) \leq K$ the approximation becomes arbitrarily accurate.

proof The condition in lemma 1 can be proved.

In lemma 3 it is essentially the central limit theorem that is doing the apporximation on its own. In practice however the normal approximation can be quite poor if a few relatively large eigenvalues are present but it is precisely under these conditions that $(\lambda_{k+1,n}^2/\sum\lambda_{j,n}^2)$ is rapidly decreasing in k, implying that Q_2 approaches normality much faster than Q.

We note that if $\liminf_{n\to\infty} \sum \lambda_{j,n}^2 < \infty$, then a bounded sequence $k(n)$ may not be sufficient to ensure convergence in distribution of Q^* to Q. As an example take e.g. the sequence $\lambda_{j,n} = \frac{1}{j}$, then Q_1 will never completely dominate Q_2 for bounded $k(n)$. Fortunately the theorem in the main text guarantees that for given k and n the approximation has a sort of minimum qualities not depending on the specific set $\{\lambda_{j,n}\}$. In this particular example the approximation will be quite good even for moderate values of k because both Q_1 will quickly outweigh Q_2 and Q_2 will approach normality if k grows.

TABEL 1A

n=40 δ=0

p_{targ}	k	0	0	0	0	0	0	0	0	0	0
0.005		-49	-50	-46	24	38	19	-50	-49	-50	-50
0.010		-90	-98	-82	27	40	20	-99	-91	-100	-100
0.025		-171	-215	-147	18	26	14	-222	-176	-250	-241
0.050		-218	-308	-179	-7	-14	-5	-323	-228	-491	-402
0.100		-173	-280	-135	-57	-94	-43	-292	-185	-726	-451
0.900		-62	-55	-60	57	94	43	-61	-61	15	-14
0.950		-125	-145	-112	7	14	5	-151	-127	-159	-148
0.975		-132	-163	-115	-18	-26	-14	-168	-136	-217	-188
0.990		-108	-142	-92	-27	-40	-20	-145	-112	-215	-178
0.995		-85	-116	-71	-24	-38	-19	-119	-89	-192	-154

n=80 δ=0

p_{targ}	k	0	0	0	0	0	0	0	0	0	0
0.005		-42	-48	-38	14	25	10	-50	-48	-50	-50
0.010		-72	-88	-62	15	27	10	-99	-89	-100	-99
0.025		-122	-168	-101	10	17	7	-222	-171	-250	-232
0.050		-145	-218	-116	-4	-10	-3	-322	-220	-486	-374
0.100		-108	-181	-83	-31	-60	-22	-292	-179	-708	-411
0.900		-53	-54	-48	31	60	22	-61	-59	16	-12
0.950		-97	-122	-84	4	10	3	-151	-125	-158	-142
0.975		-99	-132	-84	-10	-17	-7	-168	-133	-216	-181
0.990		-79	-111	-65	-15	-27	-10	-145	-110	-214	-171
0.995		-60	-89	-49	-14	-25	-10	-119	-87	-191	-148

n=160 δ=0

p_{targ}	k	0	0	0	0	0	0	0	0	0	0
0.005		-33	-42	-28	7	15	5	-50	-48	-50	-50
0.010		-54	-72	-45	8	17	5	-99	-89	-100	-99
0.025		-85	-125	-69	5	10	3	-222	-171	-250	-226
0.050		-97	-152	-77	-2	-6	-2	-322	-219	-484	-355
0.100		-69	-119	-53	-16	-36	-11	-292	-179	-700	-387
0.900		-42	-48	-37	16	36	11	-61	-59	16	-11
0.950		-73	-98	-61	2	6	2	-151	-124	-157	-138
0.975		-73	-103	-61	-5	-10	-3	-168	-133	-215	-176
0.990		-56	-84	-46	-8	-17	-5	-145	-110	-213	-166
0.995		-42	-65	-34	-7	-15	-5	-119	-87	-190	-144

$(p_{actual} - p_{target}) \times 10000$ for sets (1)-(10)
δ=0 corresponds to normal approximation

TABLE 1B

n=40	δ=0.006	k=min(k,12)								
p_{targ}	k	10	6	12	10	7	12	6	9	3	4
0.005		-37	-40	-38	5	5	7	-41	-35	-44	-39
0.010		-58	-64	-61	7	8	9	-66	-54	-71	-61
0.025		-86	-94	-93	7	13	7	-97	-79	-102	-88
0.050		-88	-93	-101	1	13	0	-95	-80	-92	-83
0.100		-49	-44	-63	-13	0	-17	-44	-42	-24	-32
0.900		-33	-32	-39	13	15	17	-30	-30	-15	-25
0.950		-41	-30	-55	-1	4	0	-29	-35	-8	-17
0.975		-34	-22	-49	-7	0	-7	-21	-28	-4	-9
0.990		-21	-11	-33	-7	-1	-9	-11	-17	-1	-4
0.995		-13	-6	-23	-5	-1	-7	-7	-10	-1	-2

n=80	δ=0.006	k=min(k,12)								
p_{targ}	k	12	8	12	12	8	12	6	9	3	4
0.005		-33	-33	-33	6	4	6	-41	-34	-42	-35
0.010		-51	-51	-52	7	5	7	-66	-52	-67	-54
0.025		-78	-74	-81	5	6	5	-97	-76	-97	-77
0.050		-84	-75	-90	-1	1	-2	-95	-77	-88	-74
0.100		-53	-40	-61	-14	-10	-15	-44	-41	-25	-31
0.900		-35	-31	-39	14	10	15	-30	-30	-15	-25
0.950		-51	-35	-62	1	-1	2	-29	-35	-8	-17
0.975		-46	-27	-60	-5	-6	-5	-21	-28	-4	-10
0.990		-32	-16	-44	-7	-5	-7	-11	-17	-1	-4
0.995		-21	-10	-31	-6	-4	-6	-7	-10	-1	-2

n=160	δ=0.006	k=min(k,12)								
p_{targ}	k	12	10	12	12	10	12	6	9	3	4
0.005		-28	-28	-26	5	4	4	-41	-34	-41	-32
0.010		-44	-42	-41	5	5	5	-66	-52	-65	-49
0.025		-67	-61	-62	4	4	3	-97	-76	-94	-70
0.050		-74	-64	-69	-1	0	-1	-95	-77	-86	-68
0.100		-49	-37	-46	-11	-9	-9	-44	-41	-25	-30
0.900		-34	-29	-33	11	9	9	-30	-30	-15	-24
0.950		-54	-37	-54	1	0	1	-29	-35	-8	-17
0.975		-51	-31	-52	-4	-4	-3	-21	-28	-4	-10
0.990		-37	-20	-39	-5	-5	-5	-11	-17	-1	-4
0.995		-27	-12	-28	-5	-4	-4	-7	-10	-1	-2

$(P_{actual}-P_{target})\times 10000$ for sets (1)-(10)

TABEL 1C

n=40

P_{targ}	k	17	10	23	17	11	22	10	15	5	7
0.005		-21	-27	-21	0	0	2	-24	-21	-31	-29
0.010		-28	-38	-30	0	0	2	-33	-30	-44	-41
0.025		-36	-48	-40	-1	0	3	-39	-38	-51	-51
0.050		-32	-41	-38	-2	-2	0	-31	-33	-36	-42
0.100		-13	-12	-19	-6	-6	-5	-6	-13	3	-8
0.900		-13	-14	-15	0	-2	5	-10	-13	-6	-12
0.950		-14	-12	-19	-4	-6	0	-9	-14	-3	-8
0.975		-11	-8	-15	-5	-5	-3	-6	-10	-1	-4
0.990		-6	-4	-10	-3	-3	-2	-3	-6	-1	-2
0.995		-4	-2	-6	-2	-2	-2	-2	-4	0	-1

n=80

P_{targ}	k	27	15	39	27	16	38	10	15	5	8
0.005		-18	-21	-18	2	1	2	-24	-20	-29	-23
0.010		-25	-30	-27	3	2	2	-33	-29	-41	-32
0.025		-34	-38	-37	3	2	2	-38	-36	-49	-39
0.050		-33	-35	-38	2	1	0	-31	-33	-35	-33
0.100		-17	-14	-21	-3	-3	-5	-6	-13	2	-7
0.900		-15	-15	-17	5	3	5	-10	-13	-6	-10
0.950		-18	-15	-24	2	-1	0	-9	-14	-3	-7
0.975		-15	-11	-21	0	-2	-2	-6	-11	-1	-4
0.990		-10	-6	-14	-1	-2	-2	-3	-6	0	-2
0.995		-6	-4	-9	-1	-1	-2	-2	-4	0	-1

n=160

P_{targ}	k	42	22	63	42	23	61	10	15	5	8
0.005		-15	-17	-16	2	1	2	-24	-20	-28	-21
0.010		-22	-24	-24	2	1	2	-33	-29	-40	-29
0.025		-31	-32	-33	1	0	1	-38	-36	-47	-36
0.050		-31	-31	-35	0	-1	-1	-31	-33	-35	-31
0.100		-18	-15	-21	-4	-4	-4	-6	-13	1	-8
0.900		-15	-15	-17	4	2	4	-10	-13	-6	-10
0.950		-21	-17	-25	0	-1	0	-9	-14	-3	-7
0.975		-19	-13	-24	-1	-3	-2	-6	-11	-1	-4
0.990		-13	-8	-16	-2	-2	-2	-3	-6	0	-2
0.995		-8	-5	-11	-2	-2	-2	-2	-4	0	-1

δ=0.001 $(P_{actual} - P_{target}) \times 10000$

TABLE 2A

δ	p_{targ}	n=40 e_{max}	k	set	n=80 e_{max}	k	set	n=160 e_{max}	k	set
0.0300	0.050	0.0161	4	1	0.0154	4	7	0.0154	4	7
0.0200	0.050	0.0155	4	7	0.0154	4	7	0.0154	4	7
0.0100	0.050	0.0123	5	7	0.0122	5	7	0.0122	5	7
0.0080	0.025[a]	0.0113	5	2	0.0097	6	7	0.0097	6	7
0.0060	0.025	0.0097	6	7	0.0097	6	7	0.0097	6	7
0.0040	0.025	0.0088	4	10	0.0079	7	7	0.0079	7	7
0.0020	0.025	0.0071	4	9	0.0067	4	9	0.0065	4	9
0.0010	0.025	0.0051	5	9	0.0049	5	9	0.0047	5	9
0.0008	0.025	0.0051	5	9	0.0049	5	9	0.0047	5	9
0.0006	0.025	0.0051	5	9	0.0049	5	9	0.0047	5	9
0.0004	0.025	0.0038	6	9	0.0036	6	9	0.0035	6	9
0.0002	0.025	0.0029	7	9	0.0028	7	9	0.0027	7	9
0.0001	0.010	0.0019	9	9	0.0018	9	9	0.0017	9	9

[a] 0.050 for n=40 maximum errors e_{max}

TABLE 2B

k	n=40 e_{max}	δ	set	n=80 e_{max}	δ	set	n=160 e_{max}	δ	set
6	0.0140	0.0165	3	0.0113	0.0115	8	0.0113	0.0115	8
7	0.0133	0.0141	3	0.0107	0.0135	1	0.0100	0.0084	8
8	0.0127	0.0121	3	0.0103	0.0117	1	0.0087	0.0062	8
9	0.0120	0.0103	3	0.0098	0.0102	1	0.0080	0.0096	1
10	0.0114	0.0088	3	0.0095	0.0093	3	0.0078	0.0089	1
11	0.0107	0.0075	3	0.0093	0.0087	3	0.0076	0.0083	1
12	0.0101	0.0064	3	0.0090	0.0081	3	0.0074	0.0077	1
13	0.0095	0.0054	3	0.0088	0.0075	3	0.0072	0.0071	1
14	0.0088	0.0046	3	0.0086	0.0070	3	0.0070	0.0066	1
16	0.0076	0.0033	3	0.0082	0.0060	3	0.0067	0.0057	1
18	0.0064	0.0023	3	0.0078	0.0052	3	0.0065	0.0049	3
20	0.0054	0.0016	3	0.0074	0.0045	3	0.0063	0.0046	3

maximum errors e_{max} always at p_{target}=0.05

TABLE 3

34 iterations		30 iterations		39 iterations	
λ_i	$\lambda_i - \tilde{\lambda}_i$	λ_i	$\lambda_i - \tilde{\lambda}_i$	λ_i	$\lambda_i - \tilde{\lambda}_i$
0.368435	0.0000008	0.138388	0.001798	-0.191243	-0.000301
0.310321	0.0000520	0.137991	0.001896	0.191243	0.002490
0.276326	0.0034475	0.137589	0.001497	-0.186402	-0.000983
0.252207	0.0138658	0.137181	0.007521	0.186402	0.000612
0.233498	-0.0037015	0.136769	0.008028	-0.181560	-0.005685
0.218212	0.0105486	0.136350	0.008010	0.181560	0.004358
0.205288	0.0000990	0.135926	0.019393	-0.176719	-0.007053
0.194092	-0.0109916	0.135497	0.021067	0.176719	-0.000209
0.184217	0.0129975	0.135061	0.021893	0.171877	0.004311
0.175384	0.0067024	0.134619	0.033813	-0.171877	-0.005306
0.167393	0.0211043	0.134171	0.036207	-0.167035	-0.011238
0.160098	0.0338830	0.133717	0.037429	0.167035	0.007160

dominating eigenvalues λ_i for sets (2), (3) and (4) for n=80
difference $\lambda_i - \tilde{\lambda}_i$ between exact and approximated values

TABLE 4A

p_{targ}	k=0	k=4		k=8		k=12	
		λ_j	$\tilde{\lambda}_j$	λ_j	$\tilde{\lambda}_j$	λ_j	$\tilde{\lambda}_j$
			n=40 set (1)				
0.005	0.0001	0.0004	0.0004	0.0010	0.0008	0.0017	0.0013
0.010	0.0010	0.0020	0.0019	0.0034	0.0031	0.0051	0.0042
0.025	0.0079	0.0112	0.0108	0.0147	0.0141	0.0181	0.0164
0.050	0.0282	0.0339	0.0333	0.0390	0.0381	0.0431	0.0410
0.100	0.0827	0.0887	0.0881	0.0933	0.0925	0.0965	0.0947
0.900	0.8938	0.8944	0.8944	0.8959	0.8958	0.8974	0.8969
0.950	0.9375	0.9414	0.9411	0.9447	0.9442	0.9469	0.9459
0.975	0.9618	0.9671	0.9666	0.9705	0.9699	0.9725	0.9715
0.990	0.9792	0.9845	0.9840	0.9871	0.9867	0.9885	0.9877
0.995	0.9865	0.9912	0.9908	0.9931	0.9928	0.9940	0.9935
			n=80 set (1)				
0.005	0.0008	0.0011	0.0010	0.0014	0.0013	0.0017	0.0015
0.010	0.0028	0.0035	0.0034	0.0042	0.0040	0.0049	0.0044
0.025	0.0128	0.0143	0.0142	0.0158	0.0155	0.0172	0.0163
0.050	0.0355	0.0377	0.0375	0.0397	0.0394	0.0416	0.0404
0.100	0.0892	0.0913	0.0911	0.0931	0.0928	0.0947	0.0936
0.900	0.8947	0.8952	0.8952	0.8958	0.8957	0.8965	0.8961
0.950	0.9403	0.9420	0.9419	0.9436	0.9433	0.9449	0.9441
0.975	0.9651	0.9672	0.9670	0.9690	0.9686	0.9704	0.9694
0.990	0.9821	0.9842	0.9840	0.9857	0.9854	0.9868	0.9861
0.995	0.9890	0.9908	0.9906	0.9920	0.9918	0.9929	0.9923
			n=160 set (1)				
0.005	0.0017	0.0018	0.0018	0.0020	0.0020	0.0022	0.0020
0.010	0.0046	0.0050	0.0049	0.0053	0.0052	0.0056	0.0053
0.025	0.0165	0.0171	0.0170	0.0177	0.0175	0.0183	0.0178
0.050	0.0403	0.0411	0.0410	0.0419	0.0416	0.0426	0.0420
0.100	0.0931	0.0938	0.0937	0.0944	0.0942	0.0951	0.0945
0.900	0.8958	0.8961	0.8960	0.8963	0.8963	0.8966	0.8964
0.950	0.9427	0.9434	0.9433	0.9440	0.9438	0.9446	0.9441
0.975	0.9677	0.9685	0.9684	0.9692	0.9690	0.9699	0.9693
0.990	0.9844	0.9851	0.9850	0.9857	0.9855	0.9863	0.9858
0.995	0.9908	0.9914	0.9913	0.9919	0.9917	0.9923	0.9920

tailprobabilities p_{actual} for k exact and approximated λ's

TABLE 4B

n=40 set (2)

p_{targ}	k=0	k=4		k=8		k=12	
		λ_j	$\tilde{\lambda}_j$	λ_j	$\tilde{\lambda}_j$	λ_j	$\tilde{\lambda}_j$
0.005	0.0000	0.0005	0.0004	0.0016	0.0015	0.0029	0.0024
0.010	0.0002	0.0022	0.0022	0.0050	0.0046	0.0072	0.0064
0.025	0.0035	0.0124	0.0122	0.0182	0.0176	0.0217	0.0204
0.050	0.0192	0.0363	0.0360	0.0438	0.0430	0.0474	0.0460
0.100	0.0720	0.0920	0.0918	0.0977	0.0970	0.0995	0.0985
0.900	0.8945	0.8953	0.8952	0.8979	0.8978	0.8991	0.8988
0.950	0.9355	0.9450	0.9449	0.9481	0.9478	0.9492	0.9488
0.975	0.9587	0.9713	0.9712	0.9737	0.9734	0.9745	0.9741
0.990	0.9758	0.9880	0.9879	0.9893	0.9892	0.9897	0.9895
0.995	0.9834	0.9938	0.9938	0.9946	0.9945	0.9948	0.9947
n=80 set (2)							
0.005	0.0002	0.0009	0.0009	0.0017	0.0016	0.0024	0.0021
0.010	0.0012	0.0032	0.0032	0.0049	0.0048	0.0062	0.0057
0.025	0.0082	0.0141	0.0141	0.0176	0.0174	0.0199	0.0190
0.050	0.0282	0.0379	0.0379	0.0425	0.0422	0.0452	0.0441
0.100	0.0819	0.0922	0.0922	0.0960	0.0957	0.0978	0.0970
0.900	0.8946	0.8952	0.8952	0.8969	0.8968	0.8980	0.8976
0.950	0.9378	0.9439	0.9439	0.9465	0.9464	0.9479	0.9474
0.975	0.9618	0.9699	0.9699	0.9723	0.9722	0.9734	0.9730
0.990	0.9789	0.9869	0.9869	0.9884	0.9883	0.9891	0.9888
0.995	0.9861	0.9931	0.9930	0.9940	0.9940	0.9945	0.9943
n=160 set (2)							
0.005	0.0008	0.0015	0.0015	0.0020	0.0020	0.0025	0.0023
0.010	0.0028	0.0043	0.0043	0.0054	0.0053	0.0062	0.0060
0.025	0.0125	0.0161	0.0161	0.0181	0.0180	0.0195	0.0191
0.050	0.0348	0.0401	0.0400	0.0427	0.0425	0.0444	0.0439
0.100	0.0881	0.0933	0.0933	0.0955	0.0954	0.0968	0.0964
0.900	0.8952	0.8958	0.8958	0.8967	0.8966	0.8974	0.8973
0.950	0.9402	0.9438	0.9438	0.9457	0.9455	0.9468	0.9465
0.975	0.9647	0.9694	0.9694	0.9713	0.9712	0.9723	0.9720
0.990	0.9816	0.9862	0.9862	0.9876	0.9875	0.9884	0.9881
0.995	0.9885	0.9925	0.9925	0.9935	0.9934	0.9940	0.9938

tailprobabilities p_{actual} for k exact and approximated λ's

TABLE 4C

n=80 set (7)

p_{targ}	k=0	k=4		k=8		k=12	
		λ_j	$\tilde{\lambda}_j$	λ_j	$\tilde{\lambda}_j$	λ_j	$\tilde{\lambda}_j$
0.005	0.0000	0.0003	0.0003	0.0017	0.0016	0.0034	0.0028
0.010	0.0001	0.0017	0.0017	0.0052	0.0050	0.0079	0.0070
0.025	0.0028	0.0110	0.0110	0.0187	0.0183	0.0227	0.0214
0.050	0.0178	0.0346	0.0345	0.0445	0.0440	0.0483	0.0470
0.100	0.0708	0.0907	0.0907	0.0982	0.0978	0.0999	0.0991
0.900	0.8939	0.8951	0.8951	0.8983	0.8982	0.8994	0.8991
0.950	0.9349	0.9446	0.9446	0.9484	0.9483	0.9495	0.9491
0.975	0.9582	0.9709	0.9709	0.9739	0.9738	0.9747	0.9744
0.990	0.9755	0.9877	0.9877	0.9894	0.9893	0.9898	0.9897
0.995	0.9831	0.9936	0.9936	0.9947	0.9946	0.9949	0.9948
n=80 set (8)							
0.005	0.0002	0.0006	0.0006	0.0014	0.0014	0.0023	0.0020
0.010	0.0011	0.0025	0.0025	0.0043	0.0042	0.0061	0.0054
0.025	0.0079	0.0125	0.0124	0.0166	0.0164	0.0197	0.0185
0.050	0.0280	0.0356	0.0356	0.0412	0.0410	0.0449	0.0435
0.100	0.0821	0.0902	0.0901	0.0951	0.0949	0.0977	0.0966
0.900	0.8941	0.8948	0.8948	0.8966	0.8966	0.8980	0.8976
0.950	0.9375	0.9426	0.9426	0.9460	0.9459	0.9478	0.9472
0.975	0.9617	0.9684	0.9684	0.9717	0.9716	0.9733	0.9727
0.990	0.9790	0.9857	0.9856	0.9880	0.9879	0.9890	0.9886
0.995	0.9863	0.9921	0.9921	0.9937	0.9937	0.9944	0.9941
n=80 set (9)							
0.005	0.0000	0.0015	0.0015	0.0033	0.0032	0.0041	0.0039
0.010	0.0000	0.0048	0.0048	0.0078	0.0078	0.0089	0.0087
0.025	0.0000	0.0183	0.0183	0.0228	0.0227	0.0240	0.0237
0.050	0.0014	0.0446	0.0446	0.0488	0.0487	0.0495	0.0493
0.100	0.0292	0.0994	0.0994	0.1006	0.1005	0.1004	0.1004
0.900	0.9016	0.8991	0.8991	0.8998	0.8997	0.8999	0.8999
0.950	0.9342	0.9495	0.9495	0.9499	0.9499	0.9499	0.9499
0.975	0.9534	0.9748	0.9748	0.9749	0.9749	0.9750	0.9750
0.990	0.9686	0.9899	0.9899	0.9900	0.9900	0.9900	0.9900
0.995	0.9759	0.9950	0.9950	0.9950	0.9950	0.9950	0.9950
n=80 set (10)							
0.005	0.0000	0.0015	0.0015	0.0027	0.0025	0.0034	0.0030
0.010	0.0001	0.0046	0.0046	0.0068	0.0064	0.0079	0.0072
0.025	0.0018	0.0173	0.0173	0.0211	0.0203	0.0226	0.0216
0.050	0.0126	0.0426	0.0426	0.0467	0.0458	0.0481	0.0470
0.100	0.0589	0.0969	0.0969	0.0993	0.0986	0.0998	0.0991
0.900	0.8988	0.8975	0.8975	0.8990	0.8987	0.8994	0.8991
0.950	0.9358	0.9483	0.9483	0.9493	0.9490	0.9496	0.9493
0.975	0.9569	0.9740	0.9740	0.9746	0.9745	0.9748	0.9746
0.990	0.9729	0.9896	0.9896	0.9898	0.9898	0.9899	0.9898
0.995	0.9802	0.9948	0.9948	0.9949	0.9949	0.9950	0.9949

tailprobabilities p_{actual} for k exact and approximated λ's